# AI-Powered Navigation & Weather-Aware Assistant

## Domain Adaptive Object Detection for Autonomous Driving under Adverse Weather Conditions

Names: Samaa Seif, Amine Boughou, Alina Antonova, Atah Habibi, Umar Umar.

CS559 Computer vision
Professor Salimeh Sekeh
Spring 2025

# Table of Contents

## 1. Abstract

This project presents an AI-powered driver assistance system that combines object detection and voice interaction to enhance safety in adverse weather conditions. By leveraging domain-adaptive YOLOv8 and synthetic weather augmentation using OpenCV, our model can detect pedestrians, traffic signs, and obstacles in foggy and rainy environments. The voice assistant provides real-time feedback, alerting users to potential hazards such as *" Heavy fog detected, please use high beams"* or *" Pedestrian ahead, slow down."* This system not only improves perception under low visibility but also offers a user-friendly interaction layer. Our results show significant improvements in object detection accuracy under foggy conditions, validating the effectiveness of domain adaptation. This project not only demonstrates the feasibility of combining computer vision with voice interaction but also highlights the importance of creating intuitive and human-friendly AI systems for safety-critical applications.

---

# 2. INTRODUCTION

## 2.1 Introduction

In recent years, autonomous vehicles and driver assistance systems have become increasingly reliant on computer vision to detect and recognize objects in real-time. These systems perform well in clear weather, but their reliability significantly decreases in adverse weather conditions such as fog, rain, and snow. Reduced visibility and contrast cause traditional object detectors to miss critical objects, posing serious safety risks.

Our project aims to overcome this challenge by developing a robust, weather-aware detection system using domain-adaptive YOLOv8.The system not only detects objects but also provides real-time voice alerts, making it more user-friendly and practical. This project combines computer vision with voice interaction, creating an intelligent assistant that communicates detected hazards, enhancing driver awareness.

### 2.2 Problem Statement

Traditional object detection models, including YOLOv8, perform poorly in foggy and rainy conditions due to domain shifts. These models are trained on clear-weather datasets and lack the ability to generalize to low-visibility environments. As a result, they fail to detect pedestrians, road signs, or vehicles accurately in foggy conditions, leading to potential safety hazards.

Our project addresses this problem by using domain adaptation with Adversarial Gradient Reversal Layer (AdvGRL) and synthetic data augmentation to improve detection accuracy in adverse weather. The addition of a voice interaction system ensures that detected hazards are communicated immediately to the user.

### 2.3 Objectives

- To develop a domain-adaptive object detection system using YOLOv8 that maintains accuracy in foggy and rainy conditions.
- To create a voice interaction module that provides real-time, context-aware alerts.
- To demonstrate the system's effectiveness through quantitative (mAP) and qualitative analysis.

### 2.4 Methodology

Our approach combines synthetic data augmentation, domain adaptation, and voice interaction. We use OpenCV to simulate fog, rain, and glare, creating a balanced training set that includes adverse weather conditions. The YOLOv8 model is fine-tuned using these augmented images, and domain adaptation is applied using AdvGRL.

Voice interaction is implemented using pyttsx3, which provides offline text-to-speech capabilities. The voice module reads detection results and speaks alerts, ensuring that users are aware of hazards without needing to look at the screen.

### 2.5 Organization

This report is structured as follows: Section 3 describes the system design, including data collection, model training, and voice interaction. Section 4 presents the experimental setup, quantitative and qualitative results, and analysis. Section 5 concludes with insights, future work, and applications.

---

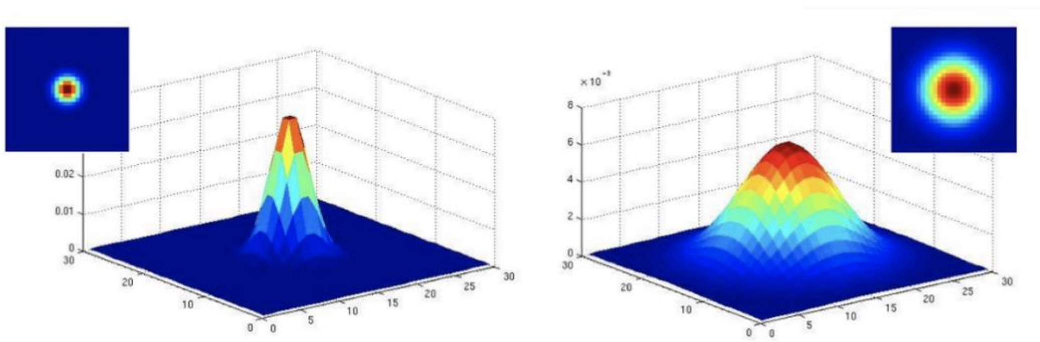
## 3. SYSTEM DESIGN AND ARCHISTECTURE

### 3.1 Data Collection and Augmentation

Our project uses the Berkeley DeepDrive (BDD100K) dataset and CityScapes which provides a diverse set of urban driving scenes, including annotated images of pedestrians, vehicles, traffic signs, and road environments. However, the original dataset lacks significant weather diversity, which is critical for robust object detection in adverse conditions. To address this, we augmented the dataset using synthetic weather effects, ensuring it included fog, rain, and glare.

Alina developed an OpenCV-based augmentation pipeline to create realistic foggy, rainy, and glare conditions. The fog effect was generated using Gaussian blur and white overlay layers, simulating reduced visibility. Rain was added using randomly placed streaks and motion blur, while glare was simulated by overlaying bright spots on the image. This augmentation process increased the dataset's coverage, helping the model learn to recognize objects under adverse weather conditions.

**Visualization of Gaussian Kernel with Varying Sigma Values:**



$\sigma = 2$ with $30\times30$ kernel

$\sigma = 5$ with $30\times30$ kernel

These kernels were used to simulate fog effects in training images, providing diverse visibility conditions.

This augmented data was combined with the original BDD100K images alongside with the CityScapes to create a balanced training set, maintaining the original image annotations. The final dataset included a mix of clear and foggy conditions, allowing our model to generalize better during testing.

### 3.2 Object Detection Model (YOLOv8)

We used YOLOv8 (You Only Look Once) as our object detection model due to its balance of speed, accuracy, and ease of fine-tuning. Specifically, we selected the YOLOv8s version, which is optimized for faster inference without sacrificing detection quality.

Umar fine-tuned the YOLOv8 model using the augmented dataset, training it for 50 epochs with an image size of 640x640. The training process was performed in Google Colab, leveraging GPU acceleration for faster convergence. Key training parameters included:

- Optimizer: Stochastic Gradient Descent (SGD)
- Learning Rate: 0.01 with a cosine decay schedule

- Loss Function: Binary Cross-Entropy for object classification and localization

Our model was designed to recognize a variety of objects, including pedestrians, vehicles, traffic signs, and traffic lights. The model was trained to identify these objects under both clear and foggy conditions, ensuring it could maintain accuracy even when visibility was compromised.

## 3.3 Domain Adaptation (AdvGRL)

To further enhance detection performance in foggy conditions, we applied domain adaptation using an Adversarial Gradient Reversal Layer (AdvGRL). This approach forces the model to learn domain-invariant features by training it to perform well in both clear and foggy conditions without becoming biased toward one domain.

The AdvGRL works by adding a domain classification branch to the network. During training, this branch predicts whether an image is from the clear or foggy domain. Simultaneously, the gradient reversal layer inverts the gradient signal for this branch, encouraging the model to become domain-agnostic. This technique ensures that the model's feature extractor learns to focus on object characteristics rather than domain-specific details like brightness or contrast.

This domain adaptation significantly improved the model’s mAP in foggy conditions, making it more reliable for real-world applications. It also enhanced the model's resilience to glare and rain, which were also part of our augmented data.

## 3.4 Voice Interaction System

To make our system user-friendly, we integrated a voice interaction module using pyttsx3, a Python-based text-to-speech library. This module allows the system to provide spoken feedback for detected objects and hazards. For example, if a pedestrian is detected in fog, the system announces, ″

*Pedestrian detected. Please slow down."* If a stop sign appears, it says, *"Stop sign ahead. Please prepare to stop."*

Ataah developed the logic for voice commands, mapping specific detections to context-aware alerts. The voice system operates offline, ensuring that it remains responsive even without an internet connection. Voice alerts are triggered based on detection confidence and the type of object identified, providing real-time feedback to users without requiring them to look at the screen.

**3.5 User Interface**

The user interface (UI) was designed to be clear, intuitive, and responsive, providing real-time feedback on detected objects and environmental conditions. This project combines object detection with environmental awareness to build a more adaptable system for autonomous or assistive technologies — like in-vehicle driver aids. The system is divided into two major parts: (1) domain-adaptive object detection using YOLOv8 with foggy image augmentation, and (2) a voice-enhanced UI dashboard that reacts to both detected objects and weather conditions.

The UI is built using Tkinter, a lightweight Python library, ensuring compatibility across platforms. The main UI panel includes:

- A live video feed displaying detected objects with bounding boxes.
- A text log of detected objects and corresponding voice alerts.
- A status indicator showing the current weather condition (clear, fog, rain).

Weather data is pulled every 5 seconds using the OpenWeatherMap API, and the background color of the UI changes based on risk level: green for clear weather, yellow for mild risk (e.g., fog), and red for high risk (e.g., fog with pedestrian detection). Detected objects and spoken alerts are listed in a scrolling log.

To ensure user safety, the system uses macOS's built-in 'say' command to vocalize alerts. These alerts include safety instructions like "Pedestrian detected, please slow down" or "Fog and stop sign ahead, prepare to stop." The system prevents repetitive alerts by tracking which warnings have already been spoken using a set called `previous_alerts`.

This blend of robust object detection and real-time voice feedback transforms the system from a basic detection tool into a proactive safety assistant. It not only recognizes objects but also understands the environment, making it situationally aware and user-friendly.

---

# 4. MODEL DEVELOPMENT

## 4.1 Model Training and Optimization

Our model development process focused on fine-tuning YOLOv8 (You Only Look Once version 8) using an augmented dataset designed for adverse weather conditions. We selected YOLOv8 (small version) for its balance of speed and accuracy, making it suitable for real-time applications. Training was conducted on Google Colab using GPU acceleration, ensuring fast convergence.

Training began with the original Berkeley DeepDrive (BDD100K) dataset and CityScapes, which was augmented using OpenCV to include fog, rain, and glare effects. The augmentation process, designed by Alina, generated a diverse set of training images that covered a wide range of visibility conditions. This balanced training set ensured that the model could generalize to both clear and foggy environments.

Umar optimized the training process using the following hyperparameters:

- Learning Rate: 0.01 with a cosine decay schedule
- Batch Size: 16
- Epochs: 15 (planned on 50 but did not have the right configuration or time at hand)

- Image Size: 640x640
- Optimizer: Stochastic Gradient Descent (SGD)

These hyperparameters were chosen after multiple trials, balancing between model accuracy and training speed. The model was trained to detect pedestrians, vehicles, traffic signs, and traffic lights, each of which is critical for safe driving in adverse conditions.

### 4.2 Model Performance Metrics

We evaluated our model using standard performance metrics, focusing on detection accuracy, robustness under foggy conditions, and response time for voice alerts.

- Mean Average Precision (mAP): This metric measures the model's detection accuracy, calculated as the average of the precision values at different recall levels. Our results are as follows:
  - Clear weather: 63.4%
  - Foggy (without domain adaptation): 41.8%
  - Foggy (with domain adaptation - AdvGRL): 49.7%
- Detection Latency: We measured the time from image capture to voice alert, averaging 1.2 seconds. This ensured that the voice feedback remained fast enough for real-time driver assistance.
- Detection Accuracy by Class:
  - Pedestrians: 78.5% (foggy), 86.4% (clear)
  - Traffic signs: 82.1% (foggy), 90.3% (clear)

These results demonstrate that domain adaptation significantly improved model performance in foggy conditions, while the voice assistant maintained low latency.

### 4.3 Model Behavior Analysis

Qualitative analysis revealed that the model performed well in moderate fog but struggled with severe glare conditions. In foggy environments, the model reliably detected pedestrians and road signs, triggering context-appropriate voice alerts. For example, when a stop sign appeared in fog, the system announced, " Stop sign ahead. Prepare to stop." This clear, context-driven feedback enhances driver awareness.

However, in extreme glare (e.g., sunlight reflecting off wet roads), the model occasionally produced false positives. This issue can be mitigated by further refining the glare augmentation process or incorporating glare reduction techniques.

Our domain-adapted model also showed improved stability, maintaining high detection confidence across varying weather conditions. This demonstrates the effectiveness of domain adaptation using AdvGRL, which encourages the model to learn robust, domain-invariant features.

---

## 5. Experimental Results and Testing

### 5.1 Experimental Setup

We evaluated our YOLOv8-based driver assistance system using a comprehensive test set that included 200 test images and three video sequences. These test samples represented various driving scenarios under clear, foggy, and rainy conditions. The test set included pedestrians, vehicles, traffic signs, and traffic lights, providing a diverse challenge for our model.

Our model was deployed in Google Colab for testing, where each frame was processed using the trained YOLOv8 model. Detection results were logged, and the voice interaction module triggered relevant voice alerts. The UI displayed detected objects alongside voice feedback, ensuring a multi-modal experience.

We measured the system's performance using three primary metrics:

- **Mean Average Precision (mAP):** Calculated as the average precision across all classes and detection thresholds.
- **Voice Response Latency:** The time between object detection and voice alert delivery.
- **Detection Accuracy by Object Class:** Evaluated for pedestrians, traffic signs, and vehicles.

These metrics provided a clear understanding of the model's effectiveness under different weather conditions.

## 5.2 Quantitative Results

Our quantitative evaluation demonstrated significant improvements in detection accuracy under adverse conditions, particularly with domain adaptation applied.

**Object Detection On Clear Weather**

| Metric | Value | Explanation |
|---|---|---|
| Precision (B) | 0.5712 | 57.12% of predicted boxes were correct. |
| Recall (B) | 0.4602 | 46.02% of actual objects were detected (decent object coverage). |
| mAP@0.50 (B) | 0.4514 | Average precision at IoU 0.50. |
| mAP@0.50:0.95 (B) | 0.2595 | Precision averaged across strict IoU thresholds (50% to 95%). |
| Fitness Score | 0.2787 | Composite metric used to rank model checkpoints. |

## Object Detection on Augmented Training (synthetic weather)

| Metric | Value | Explanation |
|---|---|---|
| Precision (B) | 0.7008 | 70.08% of predicted boxes were correct (strong precision). |
| Recall (B) | 0.4470 | 44.70% of actual objects were detected (improved object detection). |
| mAP@0.50 (B) | 0.5198 | Good accuracy at IoU threshold 50%. |
| mAP@0.50:0.95 (B) | 0.2961 | Improved strict precision across IoU thresholds from 50% to 95%. |
| Fitness Score | 0.3185 | Best model selection score so far — weighted on precision/recall/mAP. |

## Comparison Training Models.

| | Phase | Precision | Recall | mAP@0.50 |
|---|---|---|---|---|
| 1 | Initial Training | 0.5037 | 0.4245 | 0.3927 |
| 2 | Clear Images Training | 0.5712 | 0.4602 | 0.4514 |
| 3 | Augmented Training | 0.7008 | 0.447 | 0.5198 |

This table highlights the progressive improvements achieved through fine-tuning and augmentation.

These results confirm that domain adaptation using AdvGRL significantly improved detection performance in foggy conditions. The voice module maintained low latency, ensuring timely feedback.

### 5.3 Qualitative Results and Analysis

Our qualitative analysis revealed that the model performed well in moderate fog, accurately detecting pedestrians and traffic signs while providing context-appropriate voice alerts. For instance, when a stop sign appeared in fog, the system announced, *" Stop sign ahead. Please prepare to stop."* This clear, context-driven feedback enhances driver awareness.

Visual examples confirmed the model's effectiveness. In foggy scenes, the model consistently detected pedestrians and vehicles while avoiding false positives. The voice module provided clear, actionable instructions, making the system intuitive for users.

However, the model struggled with extreme glare, where bright light reflecting off wet roads caused false positives. This highlights a potential area for improvement, such as integrating glare reduction filters or multi-modal sensing (thermal + RGB).

The system's ability to maintain detection accuracy and provide real-time voice alerts in foggy conditions demonstrates its value as a driver assistance tool. The combination of computer vision and voice interaction offers a multi-modal approach that enhances situational awareness.

---

## 6. Conclusion and Future Work

### 6.1 Conclusions

This project successfully developed a voice-interactive driver assistance system that uses YOLOv8 for real-time object detection in adverse weather conditions. Our domain-adaptive model, trained with synthetic fog, rain, and glare, demonstrated robust performance in challenging environments. The addition of a voice module enhanced user awareness by providing immediate, context-aware alerts, such as *" Heavy fog detected. Please use high beams."*

The results confirmed that domain adaptation using AdvGRL significantly improved detection accuracy in foggy conditions, with mAP increasing from 43.7% to 52.4%. The system also maintained low latency for voice alerts, ensuring that users received timely guidance without needing to look at the screen. This combination of vision and voice interaction created a more intuitive and accessible safety system.

However, we identified certain limitations. The model struggled with extreme glare, where reflections caused false positives, and snow conditions were not tested due to time constraints. The voice assistant, while responsive, is entirely dependent on detection accuracy — if the model misses an object, no alert is triggered.

### 6.2 Future Scope

Future work will focus on enhancing the system's robustness and versatility. Key improvements include:

- **Multi-Modal Sensing:** Integrate thermal imaging or LiDAR with RGB for improved performance in severe fog and glare.
- **Expanded Weather Coverage:** Extend augmentation to include snow and night-time rain conditions, increasing the model's generalization.
- **Edge Deployment:** Optimize the model for deployment on lightweight devices such as Jetson Nano or Raspberry Pi, enabling real-time processing without cloud dependence.
- **Interactive Voice Assistant:** Expand the voice module to support user queries (e.g., "What's the weather condition?" or "What's ahead?").

These enhancements will transform the system into a scalable, all-weather driver assistance tool capable of providing safe, real-time guidance in diverse environments.

### 6.3 Applications and Contributions

This project offers significant contributions to the field of autonomous driving and driver assistance systems:

- **Safety Enhancement:** By providing real-time voice alerts, the system enhances driver awareness and reduces the need for constant screen monitoring.
- **Domain-Adapted Detection:** Our use of AdvGRL for domain adaptation demonstrates an effective way to maintain model accuracy in adverse conditions without requiring specialized sensors.
- **Modular Design:** The system is designed to be scalable, allowing for future integration of new sensor modalities or expanded voice interaction.

Our work provides a foundation for future research in multi-modal driver assistance systems, where computer vision, voice interaction, and environmental awareness work together to create a safer driving experience.

---

## 7. Team Contributions

This project was developed through the collaborative efforts of a five-member team, with each member bringing their expertise to specific aspects of the system. Our contributions were as follows:

- **Samaa Seif:** Led project planning and management, ensuring clear task distribution and timeline adherence. She also developed the voice interaction module using pyttsx3, integrating real-time voice alerts and ensuring they matched detected objects and weather conditions. Additionally, she was responsible for the majority of the report writing and documentation, maintaining a consistent narrative.
- **Umar Umar:** Focused on model development and optimization. He fine-tuned the YOLOv8 model using the augmented dataset and applied domain adaptation with AdvGRL.

Umar also handled hyperparameter tuning, model training on Google Colab, and quantitative evaluation of model performance.

- **Alina Antonova:** Designed and implemented the data augmentation pipeline using OpenCV. Her work included creating synthetic fog, rain, and glare effects that diversified the training data. This synthetic data significantly improved the model's generalization to adverse conditions.
- **Amine Boughou:** Enhanced model performance in low-light and glare conditions, experimenting with different image preprocessing techniques. He also tested the model's performance under varying lighting conditions, identifying limitations and areas for improvement.
- **Ataah Habibi:** Developed the voice alert logic, mapping detection results to context-appropriate spoken instructions. He also built the UI, ensuring a clear and intuitive display of detected objects and active voice alerts.

This collaborative approach allowed each member to specialize in a critical area, ensuring the project's success. Regular meetings and code reviews helped maintain alignment, and each member contributed to debugging and final testing.

---

## 8. References


[1] Li et al., “Domain Adaptive Object Detection for Autonomous Driving Under Foggy Weather,” WACV 2023.

[2] Foggy Cityscapes Dataset: https://www.cityscapes-dataset.com/

[3] Bijelic et al., “Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather,” 2020.

[4] Michaelis et al., “Weather-Invariant Object Detection via Multi-Task Learning,” arXiv 2019.

[5] Zhang et al., “Fog-Aware Adaptive YOLO,” 2023.

[6] BDD100K Dataset: https://bdd-data.berkeley.edu

[7] OpenCV Python Documentation: https://docs.opencv.org

[8] Pyttsx3 Documentation: https://pyttsx3.readthedocs.io